

1. Opening remarks and introductions

The Chairman, Don Tolmie of Los Alamos National Laboratory, opened this HIPPI-6400 meeting and thanked Charles Brill and AMP for hosting this meeting. This group is constituted as both the HIPPI special working group (SWG) under X3T11, and the HIPPI Networking Forum (HNF) - Technical Committee (TC).

Don then lead a round of introductions. The list of attendees is at the end of these minutes.

2. Review / modify the draft agenda

The draft agendas were available on the web prior to the meeting. Hard copies were available at the meeting. James Hoffman of Los Alamos volunteered to take the meeting minutes.

The group inserted item 4.8: Document split, to decide whether the document should be split up to ease readability and future upgrades.

3. Review minutes of previous meetings

3.1 September 11-12, 1996, Albuquerque, NM

The Albuquerque minutes were reviewed and approved with no the changes.

3.2 Review action items from Albuquerque meeting

1. James Hoffman to verify CRC error protection results. (Done)
2. Greg Chesson and SGI to provide text on what actions occur as a result of a Shutdown. (Done)
3. Greg Chesson to provide ARP text for inclusion in HIPPI-6400-SC and specify affects on bridging. (Carryover)
4. Greg Chesson to draft initial text for bridging. (Done, merged with above)
5. Hansel Collins to check the 0.6 ns (20-80%) rise and fall time values. (Carryover)
6. Michael McGowen to begin VC Message size restrictions discussion via email. (Carryover)
7. Greg Chesson to review counter size of SuMAC Retransmission_Error counter and the need for

both contiguous retransmission and total retransmission error counters. (Carryover)

8. Dave Parry to review the changed values for stall timeout and credit timeout. (Carryover)
9. Dave Parry or Jim Davis to review the new text of 10.1.2: Check Received TSEQ. (Done)
10. Michael McGowen to look at bridging and address self discovery concerns with the new MAC Header and start a discussion on email for any unresolved issues. (Carryover)
11. Greg Chesson to review port mapping for HIPPI-6400 Scheduled Transfers. (Done)
12. Greg Chesson to look into IP over HIPPI-6400, focusing on the Scheduled Header. (Done)
13. Roger Ronald to update the Admin micropacket draft. (Done)
14. Don Tolmie to update HIPPI-6400-PH Rev 0.5 with the changes agreed to at the Albuquerque meeting. (Done)

4. Review HIPPI-6400-PH changes since last meeting (reference HIPPI-6400-PH Rev 0.6)

4.1 Minor changes first and then return to unresolved issues in 4.8

Changed the length for copper and fiber implementations from "TBD" to 50 m and 200 m respectively. This is in the Abstract, Foreword, Introduction, and Scope – accepted with rewording of the second to last sentence.

In the Introduction, 8th bullet, changed "error free" to "reliable" – accepted.

Added ISO/IEC 8802-2 as a normative reference – accepted.

Added "...and low processing overhead..." to the definition of Scheduled Transfer in 3.1.15 – accepted.

In figure 2, page 4, added "1(1)" and "1(0)" to the signal lines to better define the number of lines used – accepted.

In 4.5, deleted "...and a VC is reserved for this transfer" from the description of Scheduled Transfers – accepted.

In 4.6 added the CLOCK_2 signal – accepted.

In 4.11, added a maximum distance of 1 km for an optical interface – accepted. Roger Ronald and Craig Davidson commented that the wide parallel fiber interface would only be a slight modification of the narrow interface. The group agreed to place such a specification in an informative annex.

In 6.3, consolidated the text about which micropackets are retransmitted, and rather than re-specify the actions taken when unspecified TYPES are encountered just referenced the definitive text – accepted.

In 6.3.2, deleted the last sentence about Null micropackets not being retransmitted (since it is stated in 6.3) – accepted with some rewording.

In 6.4, added note 3 about reusing TSEQ of stomped micropackets – accepted.

In 6.5, the last sentence referring to the credit counter was changed so that the credit counter runs exactly from 0 to 255 credits. Credit-counter-overflow was added as a condition to 13.3: Link Shutdown. It was commented that Note 3 of 6.5 implies an implementation and should be removed, but the intention to rapidly update credits should appear in the text.

In 11.2, last paragraph, the sentence about showing the CLOCK_2 signal skewed was added. Figures 13 and 14 were modified with the CLOCK_2 signal deliberately skewed in relation to CLOCK – accepted.

In 13.3, a link Shutdown condition dealing with credit counter overflow was added. The second to last sentence in the last paragraph was removed. The thought is that a link Shutdown will cause only the “trouble” side of the link to require “administrative action” and the other side will stay in a Power-On state until link activity is detected.

Figure 17 was modified at the top by the showing that Power-On could result in either a Reset or Initialize – accepted. Greg Chesson took an action item to have the SGI review Figure 17.

In 15.1, requiring a constant phase relationship between CLOCK and CLOCK_2 was added at the end of the first paragraph – accepted.

In 15.2, the change is the deletion of the last sentence saying that one of the CLOCK signals could be used as an inactivity monitor – accepted.

15.3 was changed by removing the requirement for 1 ns between the rising edges of the CLOCK and CLOCK_2 signals.

4.2 LCRC specification and parallel equations (pages 12-13, 42-47)

James Hoffman presented CRC error coverage results which appear on the web at: <http://www.cic-5.lanl.gov/~jamesh/hippi64/link.html>
The group requested a comparison to a single 32 bit CRC over the same data size and possibly continued testing for 7 and 8 bit errors.

In 6.6.2, in the paragraph starting "Figure 8..." added "Refer to tables 5...". Added all of the next paragraph to better describe the LLRC checking on the Destination side, including the residues for 'no errors' and 'stomped' micropackets – accepted with deletion of references to “figure 3” which may add confusion.

In 6.6.2.1 added the last words in the paragraph about the TSEQ not being 'consumed' – accepted with minor rewording.

In 6.6.2.2, changed from the syndrome to a reference to where the residue for a 'stomp' code was defined – accepted.

Figures 8 and 9 were redone to show the input data coming into the lsb instead of the msb – accepted.

In 6.6.3, next to last paragraph, added references to the tables describing the transmission sequence – accepted with deletion of references to “figure 3” which may add confusion.

Annex B.3 on the LCRC parallel implementation was completely redone to reflect the fact that the LCRC does not include 16 extra 0's. The parallel equations for the LCRC were shown in 16-bit, 64-bit, and 80-bit formats – accepted with a few suggestions.

Annex B.4 on the ECRC parallel implementation was also redone with additional text. The equations were also replaced to compensate for now feeding the data into the lsb instead of the msb – accepted with a few suggestions.

4.3 Error processing (pages 25-27, 36)

In 9.1 the timeout value (end of the last paragraph) was changed from 1 ms to 2 seconds – accepted.

In 9.4, the note was added pointing out that larger values may be needed in some environments – accepted.

In 10.1.2, the first sentence was changed by adding "...non-stopped,..." – accepted.

In 10.1.4, the unspecified TYPE text was consolidated (both < 8 and ≥ 8). Treating unspecified TYPEs ≥ 8 as Data replaced tracking the message-in-progress status. The group decided to split the intermediate and final destination requirements for unspecified TYPE handling. Intermediate nodes (switches) will treat all TYPEs ≥ 8 as Data whereas final nodes may use a message-in-progress indicator.

In 10.2.4, what had been 'TAIL' timeout was renamed to 'Stall' timeout, and the error flag was also renamed for consistency. The default value was changed from 2 seconds to 2 ms. A note was added about how this could be triggered by slow hosts – accepted.

A new subclause 10.2.5 was added to isolate the text denoting the no-error case – accepted.

Table 8 changed with the name change from "TAIL_Timeout_Error" to "Stall_Timeout_Error" – accepted.

Table 9 was changed with the new values for the marked timers – accepted.

On the HIPPI reflector, Dave Parry asked for a stronger retransmission scheme that sends two skew retraining sequences before retransmitting the buffer. It was noted that this may cause problems in the LANL/Optivision tester and was left open for further study.

4.4 Copper interface specifications (pages 37-39)

In 16.2, the changes were the addition of the electrical values, and inclusion of the power signal and its values – accepted with a change to Minimum low-level output voltage = .08 V. It was also noted that most of the specification reflect values measured at the chip but we want to specify bulkhead values. Hansel Collins was given an action item to modify the electrical values to what would be seen at the bulkhead.

In 16.3, the changes are the addition of the first value, and rewording of the other parameters – changed with new values (Minimum peak-to-peak differential input voltage = 150 mV).

In 16.4 the connector requirement for "10,000 mating cycles" was changed to "up to 10,000 mating cycles." – accepted with additional "fixes".

In 16.5, the cable specifications were changed to reflect the total cable, irrespective of length, rather than per meter specifications – accepted with one clarification (Jacket material = CL-2.2P/FT6).

Don noted that when the specifications have stabilized metric equivalencies will be added.

Table 10 with the connector pin assignments was completely redone. Table 10 was changed to a figure, but values will be left untouched until the group has had time to decide on the best configuration.

Now that the actual cables are available the group asked for EMC testing results possibly from SGI and IBM.

The group considered adding the remaining six pins as power pins to power outboard devices. Don Tolmie presented a list of advantages and disadvantages collected from the reflector. The main advantage was the capability of powering an outboard fiber module. The main disadvantage was requiring power from all network interface cards.

Craig Davidson motioned to reserve 6 pins for optional power capabilities. Bob Williard seconded the motion. The motion failed by a vote of 6 against and 2 for. Joe Parker moved and Greg Chesson seconded to reserve the remaining 6 pins. Motion passed unanimously.

4.5 Message Structure (pages 14-16)

The Message structure in clause 7 has major changes. For example: 1) the logical addresses were replaced with 48-bit IEEE MAC addresses, 2) the addresses were changed from 60-bit with a 4-bit registration authority to 48-bit only, 3) the M_flags and M_type were removed and replaced by the LLC/SNAP header. The M_len was changed to denote the number of bytes following the M_len parameter – accepted. The group debated whether M_len should cover all bytes of a Message (including the MAC Header), but decided to leave things unchanged to conform with Ethernet conventions.

In the Schedule Header the flags that were in the Op field were moved to a separate Flags field. What was in the original S_flags field was moved to this new Flags field also. The T_id was split into separate S_id and D_id fields. The whole header structure was

shortened and two opaque payload words were left at the end of the second micropacket – accepted with the clarification that the opaque payload would be used as additional ULP information for the Scheduled Header. The two unused 32-bit words at the end of the MAC Header during unscheduled transfers will be filled with zeros and payload starts in the second micropacket. Figure 10 – Message headers will be updated to show unscheduled transfer format.

In 7.1 and 7.2 various rewording changes were made. Greg Chesson took an action item to have SGI donate two more Ethertypes and register the Ethertypes with INA/Xerox.

In 7.3, it was noted that many fields serve double duty, but only one is described. The first paragraph will note the dual usage and that specific sections in the document should be reviewed for exact usage. The Op field was shrunk to 5 bits and the Flags field increased to 11 bits.

In 7.4, the Concatenate was required to be consistent for the whole Transfer. Rejecting an operation is now done by a 'Reject' flag bit rather than -1 in T_len or S_Port. The update flag was removed as the functionality requires excessive complexity at this level. The DATA_ACK_Requested bit was added again to ensure reliable data movement. The group requested additional text for the Concatenate and Persistent flags noting they are used between mutually agreeable hosts.

4.6 Scheduled Transfers (pages 16-24)

In 8.1, and 8.3, the fact that Blocks may be out of order, but Messages will be in order, was added – accepted with rewording suggestions for both sections.

Carrying the initial Key value was relegated to the Offset field rather than the T_id field since T_id was split into two 16-bit fields – accepted.

The Scheduled Transfer individual operations were updated with the new parameters – accepted with the suggestion to show the parameter being carried in the semantics (with the actual field in parentheses) for overloaded fields.

The group read through most of the Scheduled Transfer operation text and proposed many grammatical changes and clarification text. Please see the next iteration of the document to see specific changes.

A timeout based, retransmit operation mechanism has been inserted into 8.1 describing parameters contributing to an appropriate timeout value for resending an operation when respective response or acknowledgment operation is missing.

In 8.2.1, it was decided that an Ethertype (in B_num) will declare the upper layer protocol for the Transfer.

A new operation, Port_Teardown_Complete, was added to create a three-way handshake during port teardown. The two nodes will release buffers and other state without waiting for complex timeouts that a two-way handshake would necessitate.

A large block of text was removed from 8.3 and integrated into 8.1.

In 8.3.2, there is no specific mechanism to reject a transfer due to inability to use the concatenate or persistent flag or due to difficulty with the actual transfer. The reject flag will be set for all types of rejected transfers.

The current buffer tiling may require an Originating Source to cross buffer boundaries when sending a single Message to the Final Destination. The group agreed to revise the Message size requirement so that a Message can be bounded by both the Originating Source and Final Destination buffer boundaries. This does not stop an Originating Source from crossing buffer boundaries for a single Message; it just no longer requires it to cross Originating Source buffer boundaries.

The DATA_ACK operation provides two functions: block acknowledgment and Scheduled Transfer operation flow control. Both functions are performed each time the DATA_ACK is requested. The DATA_ACK can be requested by a Data operation with the DATA_ACK_Requested bit set or by an Request_ACK operation. In the DATA_ACK operation, the B_num field will contain an acknowledgment for all blocks correctly received starting from block zero to the listed block. This block number may be different than the block number in the Data message that the DATA_ACK answers. The second function is to provide a Scheduled Transfer operation flow control mechanism. The DATA_ACK will send an available operations update (in M_count) so that the sender does not overrun the operation descriptor queue on the receiver. A synchronization ID (in T_len) will be sent with the Data (or Request_ACK) operation which the receiver will return in the corresponding

DATA_ACK. The synchronization ID will match DATA_ACK to the operation that requested DATA_ACK for close calculation of the number of available descriptor queues (the C-limit (current limit) minus the number of operations that have been sent since the operation with the corresponding DATA_ACK_Requested bit).

Request_ACK (Sync) →

← DATA_ACK (Sync, C-limit, B_num)

Initially the group decided to keep the Sync field to allow a ULP to perform the operation flow control, but after further analysis the flow control system was added to the document to provide a basis to discuss the function.

Tables 3 and 4 were updated to reflect all of the changes.

4.7 New Annex C on Scheduled Transfers (pages 48-52)

The group didn't have time to read through the Schedule Transfer Annex C, but a request was made for a table summarizing the example. James Hoffman took the action item to add the example table and upgrade the Annex to match the document.

4.8 Resume discussions on items deferred in 4.1

The group reviewed the Shutdown text (13.3) and changed the last paragraph so as not to require administrative action on both ends after a shutdown. The group read Dave Parry's comments to the reflector about adding text to Shutdown to describe the higher layers of the interface (above the link layer). The group felt the current text was sufficient.

5. Splitting the document

Various people noted the document spans an extreme range of the network stack which may discourage third party mapping of the Scheduled Transfer and limit new physical layer implementations. There is a somewhat obvious split of the Scheduled Transfer, but the group was not sure that splitting the physical layer from the link layer had a natural boundary. Some opined that a different physical layer may also require a different link layer and thus the physical layer split would gain nothing. James Hoffman moved to begin splitting the Scheduled Transfer from the document but leave the physical layer split for

further consideration. The group unanimously agreed.

6. HIPPI-6400-SC

6.1 Review recent document changes (Reference Rev .45)

Running out of time and waylaid by the optical committee, Roger Ronald quickly sited a few changes to HIPPI-6400-SC, but no in depth comments were made. The group is encouraged to review the latest revision and send any comments to Roger Ronald.

6.2 Admin micropackets(Reference Rev 0.3)

The Admin micropacket draft was reviewed and a few word choice suggestions were made. The group asked for a definition of an "Element" and a "Link device". Someone noted that a vendor identification field might be necessary for devices that know how to access specific vendor information, but need to be able to query for the vendor code.

7. HIPPI-6400 MIB

Von Welch agreed to continue updating the MIB to account for any changes in the document.

8. Optical interconnect (8 am - 9 am, and 3 pm - 6 pm Tuesday)

The optical group met on Tuesday and discussed laser safety issues, electrical specifications, and criteria for the ferrule presentations. The ferrule presentations were made, and after discussion, a vote was held. Roger Ronald moved and Greg Chesson seconded to select the MT ferrule. The motion passed.

Please see the separate optical minutes for detailed information.

9. Future meeting schedule

9.1 November 6-7, Phoenix, AZ

The meeting will cover HIPPI-6400 issues, with emphasis on copper on Thursday afternoon.

Wednesday, November 6 — 1 PM - 9 PM

Thursday, November 7 — 8 AM - 5 PM

The location is on the Courtyard by Marriott at the Phoenix Airport. Chris Olson and Lockheed Martin are the host. (See the meeting announcement on the web page at <http://www.cic-5.lanl.gov/~det/> for further details.)

9.2 December 2-3, 1996, Minneapolis, MN

During the X3T11 December plenary week, the following HIPPI meetings are scheduled:

Monday, December 2 -

9 AM - 9 PM — HIPPI-6400

Tuesday, December 3 -

8 AM - 9 AM — HNF Plenary

9 AM - 11 AM — HIPPI-TC General and -6400

11 AM - 9 PM — HIPPI-6400 Optical (with connector presentations)

The location is the Embassy Suites, 79301 34th Avenue South, Bloomington, MN. Horst Truestedt and IBM are the host. (See the meeting announcement on the web page at <http://www.cic-5.lanl.gov/~det/> for further details.)

9.3 Future meeting dates and locations

The following 1997 X3T11 plenary week dates are firm (except for December which may be shifted one week later to avoid the Thanksgiving holiday). Recent changes to this list are underlined to make them easier to find.

1997 -

Jan 8-9	Interim	Phoenix, AZ	Lockheed
Feb 3-4	Plenary	San Jose, CA	Sun
Mar 5-6	Interim	<u>Phoenix, AZ</u>	Berg
Apr 7-8	Plenary	Palm Springs, CA	Brocade
May <u>7-8</u>	<u>Interim</u>	<u>Mt. View</u>	<u>SGI</u>
Jun 9-10	Plenary	Seattle, WA	Boeing
July <u>9-10</u>	<u>Interim</u>	<u>Minneapolis, MN</u>	<u>Cray</u>
Aug 4-5	Plenary	Honolulu, HI	Hitachi
Oct 6-7	Plenary	Tucson, AZ	FSI
Dec 1-2 <u>?</u>	Plenary	Orlando, FL	DPT

The 1998 schedule is less firm, but here is what is currently being considered by X3T11 for the plenary meetings. Question marks note the ones that are still in question. Hopefully HIPPI-6400 will be far enough along that we will not need interim working meetings.

1998 -

Feb 9-10	Plenary	San Diego	<u>Qlogic</u>
Apr 20-21	Plenary	Palm Springs, CA	Brocade
Jun 8-9	Plenary	<u>St. Petersburg</u> <u>Beach, FL</u>	AMP
Aug 10-11	Plenary	??	??
Oct 5-6	Plenary	Tucson, AZ (?)	FSI (?)
Dec 7-8	Plenary	<u>Ft. Lauderdale, FL</u>	<u>Adaptec</u>

10. Review action items

All of the following action items apply to HIPPI-6400.

1. Greg Chesson to provide ARP text for inclusion in HIPPI-6400-SC and specify affects on bridging.
2. Hansel Collins to check the 0.6 ns (20-80%) rise and fall time values.
3. Michael McGowen to begin VC Message size restrictions discussion via email.
4. Greg Chesson to review counter size of SuMAC Retransmission_Error counter and the need for both contiguous retransmission and total retransmission error counters.
5. Dave Parry to review the changed values for stall timeout and credit timeout.
6. Michael McGowen to look at bridging and address self discovery concerns with the new MAC Header and start a discussion on email for any unresolved issues.
7. Greg Chesson took an action item to have the SGI review Figure 17.
8. Hansel Collins to modify the electrical values to what would be seen at the bulkhead.
9. Henry Brandt to investigate grounding schemes for copper interface.
10. Greg Chesson to have SGI donate two more Ethertypes and register the Ethertypes with INA/Xerox.
11. James Hoffman to update the Scheduled Transfer discussion Annex C with decisions from the meeting.
12. Roger Ronald to update the Admin micropacket draft.

13. Don Tolmie to update HIPPI-6400-PH Rev 0.6 with the changes agreed to at the St. Petersburg meeting.

11. Adjournment

The meeting adjourned around 6:30 pm to AMP's free-food social.

12. Attendance

Terry Bittner	AMP
Kirk Bovill	AMP
Ed Cady	Berg
John Ellis	Berg Electronics
Jeff Young	Cray Research
Bob Willard	Digital Equipment Corp
Bob Pearson	Essential Communications
Randy Hardy	Harris
Francois Gaullier	Hewlett-Packard
Henry Brandt	IBM
James Hoffman	Los Alamos National Lab
Don Tolmie	Los Alamos National Lab
Dan Schwartz	Motorola
Von Welch	NCSA
Yasuo Sasaki	NTT International Corp.
Joe Parker	Optivision
Michael Leib	Pulse/Technitrol
Craig Davidson	Raytheon E-Systems
Roger Ronald	Raytheon E-Systems
Greg Chesson	Silicon Graphics
Hansel Collins	Silicon Graphics
Don Sanders	Silicon Graphics